

CUNY SPS DATA 621 - CTG5 - HW4

Betsy Rosalen, Gabrielle Bartomeo, Jeremy O'Brien, Lidiia Tronina, Rose Koh

April 24th, 2019

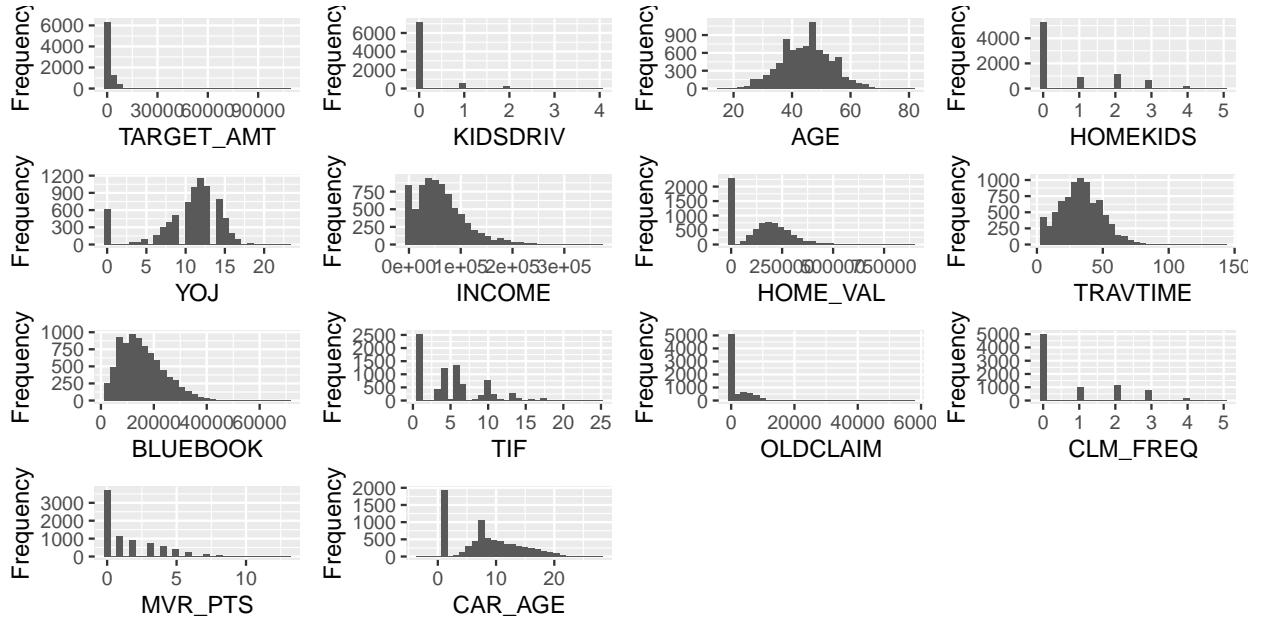
Contents

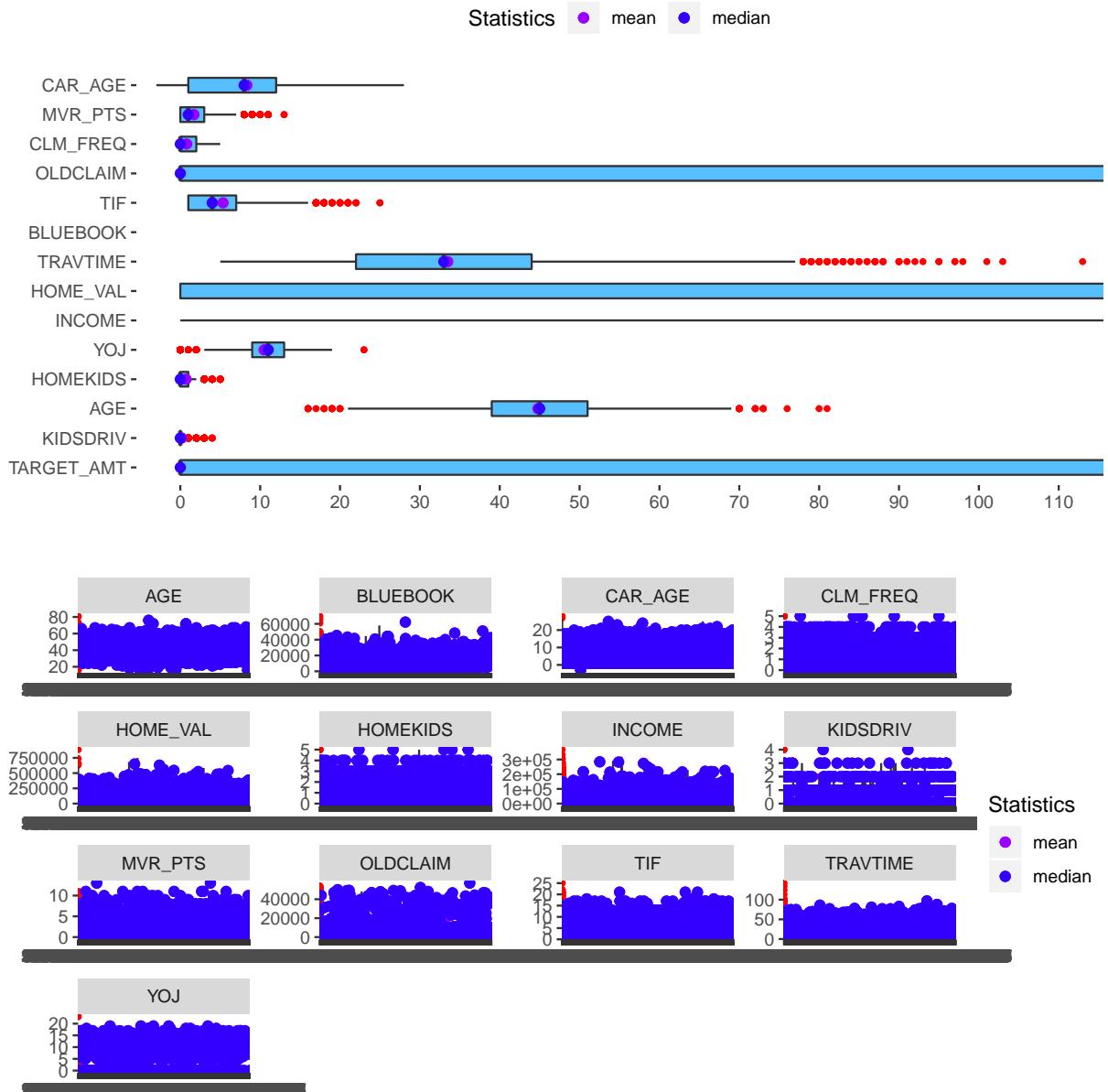
1	DATA EXPLORATION	2
2	DATA PREPARATION	7
3	BUILD MODELS	8
4	SELECT MODELS	9
5	Appendix	10

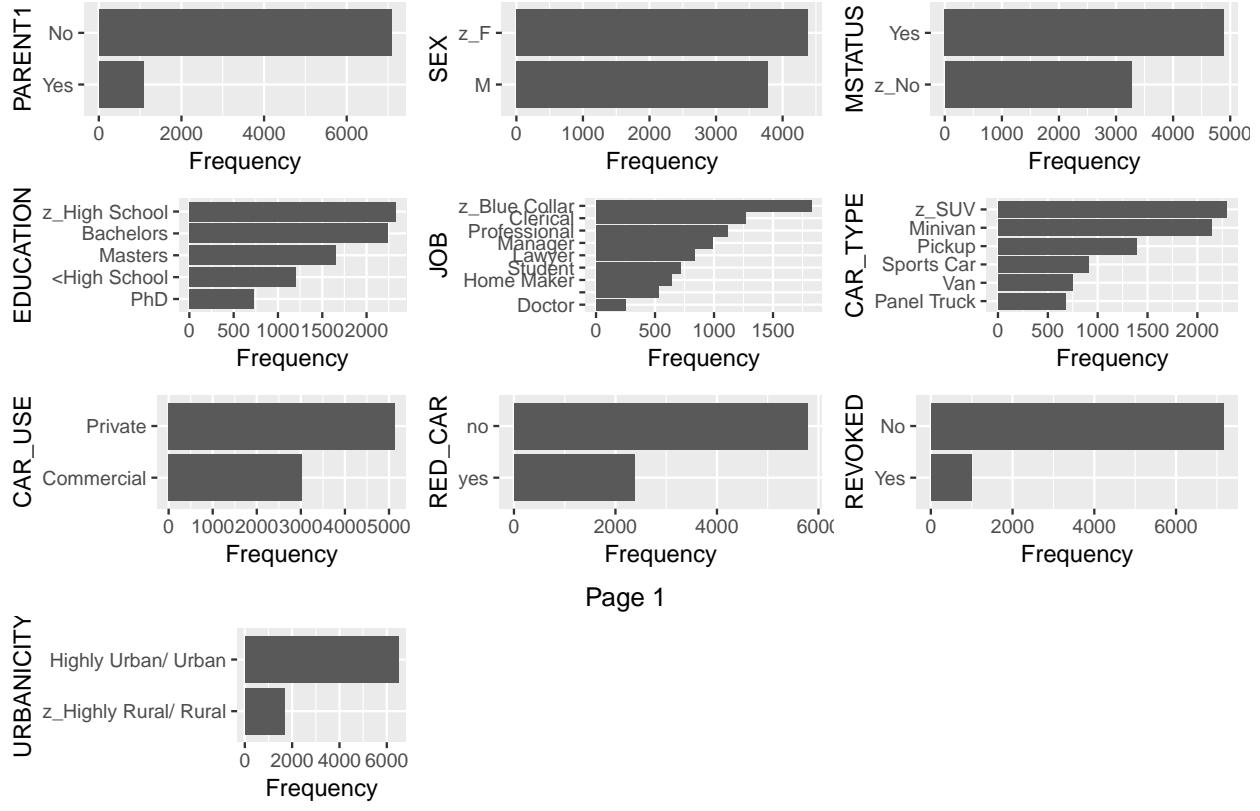
Table 1: Data Dictionary

VARIABLE	DEFINITION	TYPE
TARGET_FLAG	car crash = 1, no car crash = 0	response
TARGET_AMT	car crash cost = >0, no car crash = 0	response
AGE	driver's age - very young/old tend to be risky	numerical predictor
BLUEBOOK	\$ value of vehicle	numerical predictor
CAR_AGE	age of vehicle	numerical predictor
CAR_TYPE	type of car (6types)	categorical predictor
CAR_USE	usage of car (commercial/private)	categorical predictor
CLM_FREQ	number of claims past 5 years	numerical predictor
EDUCATION	max education level (5types)	categorical predictor
HOMEKIDS	number of children at home	numerical predictor
HOME_VAL	\$ value of home - home owners tend to drive more responsibly	numerical predictor
INCOME	\$ income - rich people tend to get into fewer crashes	numerical predictor
JOB	job category (8types, 1missing) - white collar jobs tend to be safer	categorical predictor
KIDSDRIV	number of driving children - teenagers likely get into crashes	numerical predictor
MSTATUS	marital status - married people drive more safely	categorical predictor
MVR PTS	number of traffic tickets	numerical predictor
OLDCLAIM	\$ total claims in the past 5 years	numerical predictor
PARENT1	single parent	categorical predictor
RED_CAR	a red car	categorical predictor
REVOKE	license revoked (past 7 years) - more risky driver	categorical predictor
SEX	gender - woman may have less crashes than man	categorical predictor
TIF	time in force - number of years being customer	numerical predictor
TRAVTIME	distance to work	numerical predictor
URBANCITY	urban/rural	categorical predictor
YOJ	years on job - the longer they stay more safe	numerical predictor

1 DATA EXPLORATION

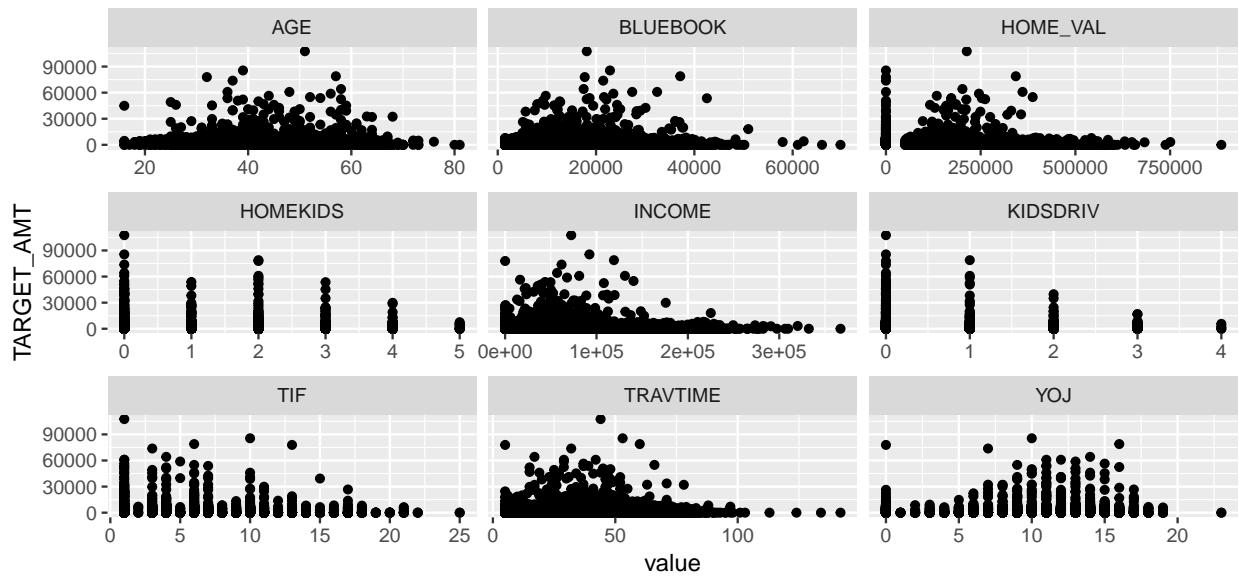




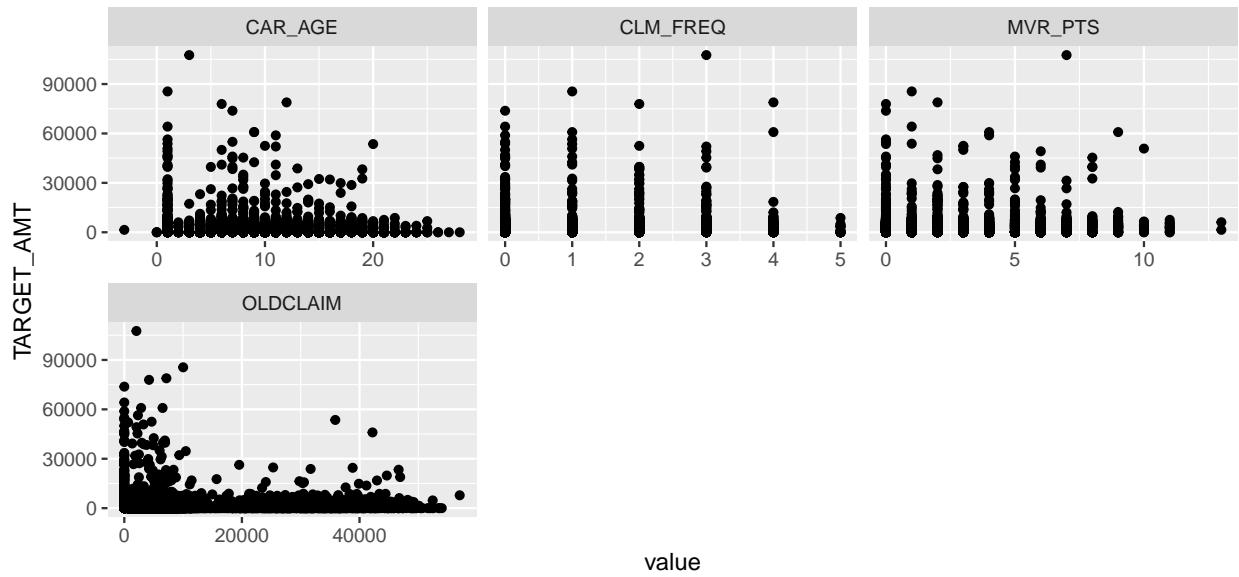


Page 1

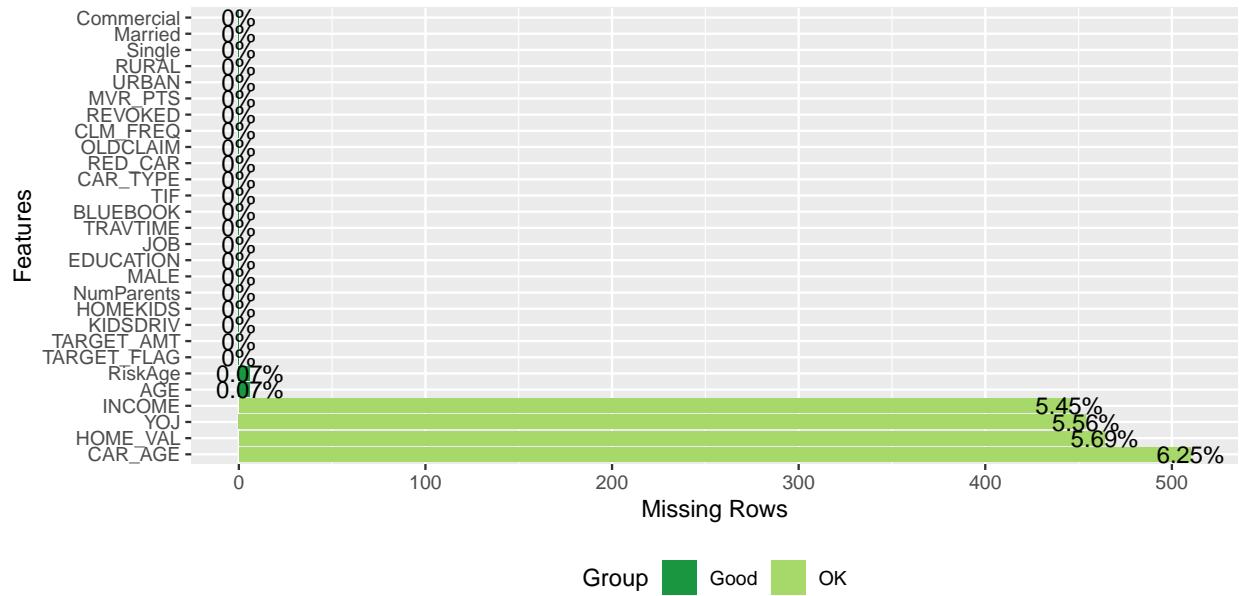
Page 2



Page 1



Page 2



2 DATA PREPARATION

3 BUILD MODELS

4 SELECT MODELS

5 Appendix

The appendix is available as script.R file in `project4_insurance` folder.

https://github.com/betsyrosalen/DATA_621_Business_Analyt_and_Data_Mining